Statistical Inference Course Project Part 1

Bamini Balaji

January 22, 2017

Part 1: Simulation

Overview

The objective is to simulate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

Simulation Code

Let's set lambda = 0.2 for all of the simulations.

First, we will simulate 1000 sets of 40 exponentials. Then, we will compute their means and store it in a vector of length 1000 named "means".

```
lambda <- 0.2
n = 40 # This indicates number of distributions to be averaged
s = 1000 # This is number of simulations
sim_data <- matrix(rexp(s*n, lambda), nrow = n, ncol = s)
means <- apply(sim_data, 2, mean)
sdev <- apply(sim_data, 2, sd)
avgmeans <- mean(means)
avgsd <- mean(sdev)
```

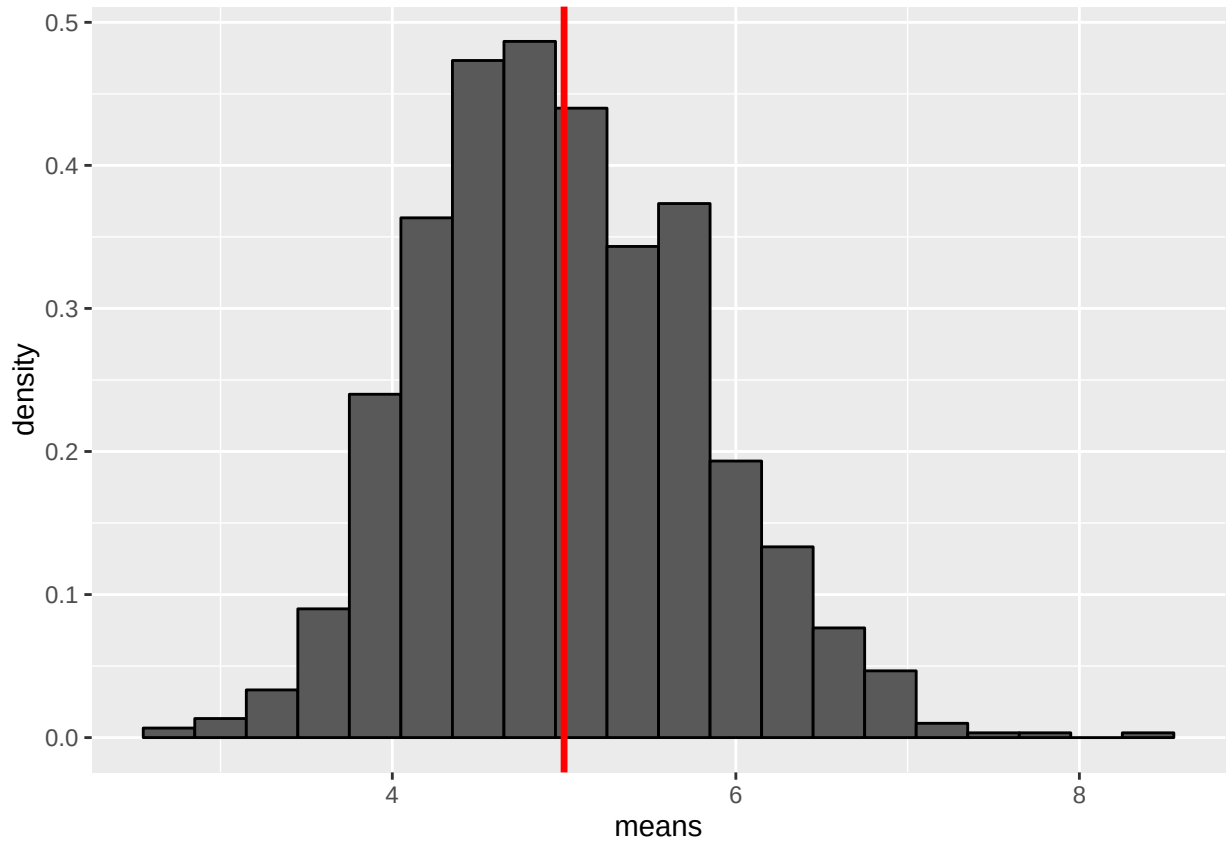
Analysis of Mean

The average of ovserved means is 4.9883204.

The theoretical mean of this distribution is 1/lambda or 5.

The following figure shows the observed distribution of means through a histogram. The expected value of the theoretical mean is shown by the red line.

This looks like a normal distribution!



Let's perform a hypothesis test with alpha of 0.05:

- H_0 : mu = 5.0
- H_a : mu != 5.0

```
t.test(means, mu = 5, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  means
## t = -0.45701, df = 999, p-value = 0.6478
## alternative hypothesis: true mean is not equal to 5
## 95 percent confidence interval:
##  4.938170 5.038471
## sample estimates:
## mean of x
##   4.98832
```

Null hypothesis is rejected if p-value is less than 0.05 (alpha).

Analysis of Variance

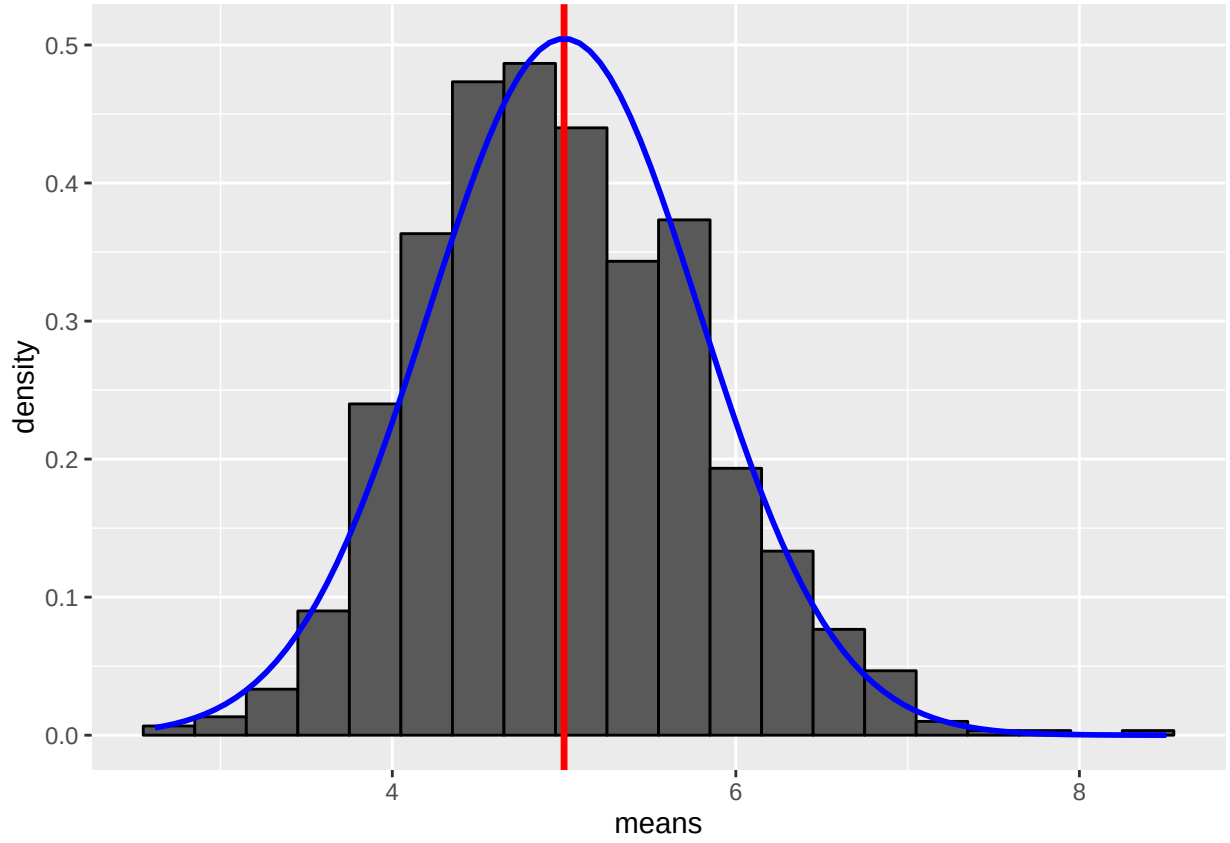
The measured average standard deviation equals 4.8717912.

The theoretical standard deviation of an exponential distribution is 1/lambda or 5.

The standard deviation of the observed distribution is theoretically sigma/sqrt(n), or 0.7905694 where n = 40 and sigma is 1/lambda.

Let's overlay a normal distribution to this figure. Z ~ N(5, 5/sqrt(40))

This distribution is illustrated using the blue curve.



Let's perform a hypothesis test with alpha of 0.05:

- H_0 : sd = 5.0
- H_a : sd != 5.0

```
t.test(sdev, mu = 5, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  sdev
## t = -3.7672, df = 999, p-value = 0.0001748
## alternative hypothesis: true mean is not equal to 5
## 95 percent confidence interval:
##  4.805007 4.938576
## sample estimates:
## mean of x
##  4.871791
```

Null hypothesis is rejected if p-value is less than 0.05 (alpha).